MicroBooNE Offline Tools Update

Eric/Herb

9 May, 2013

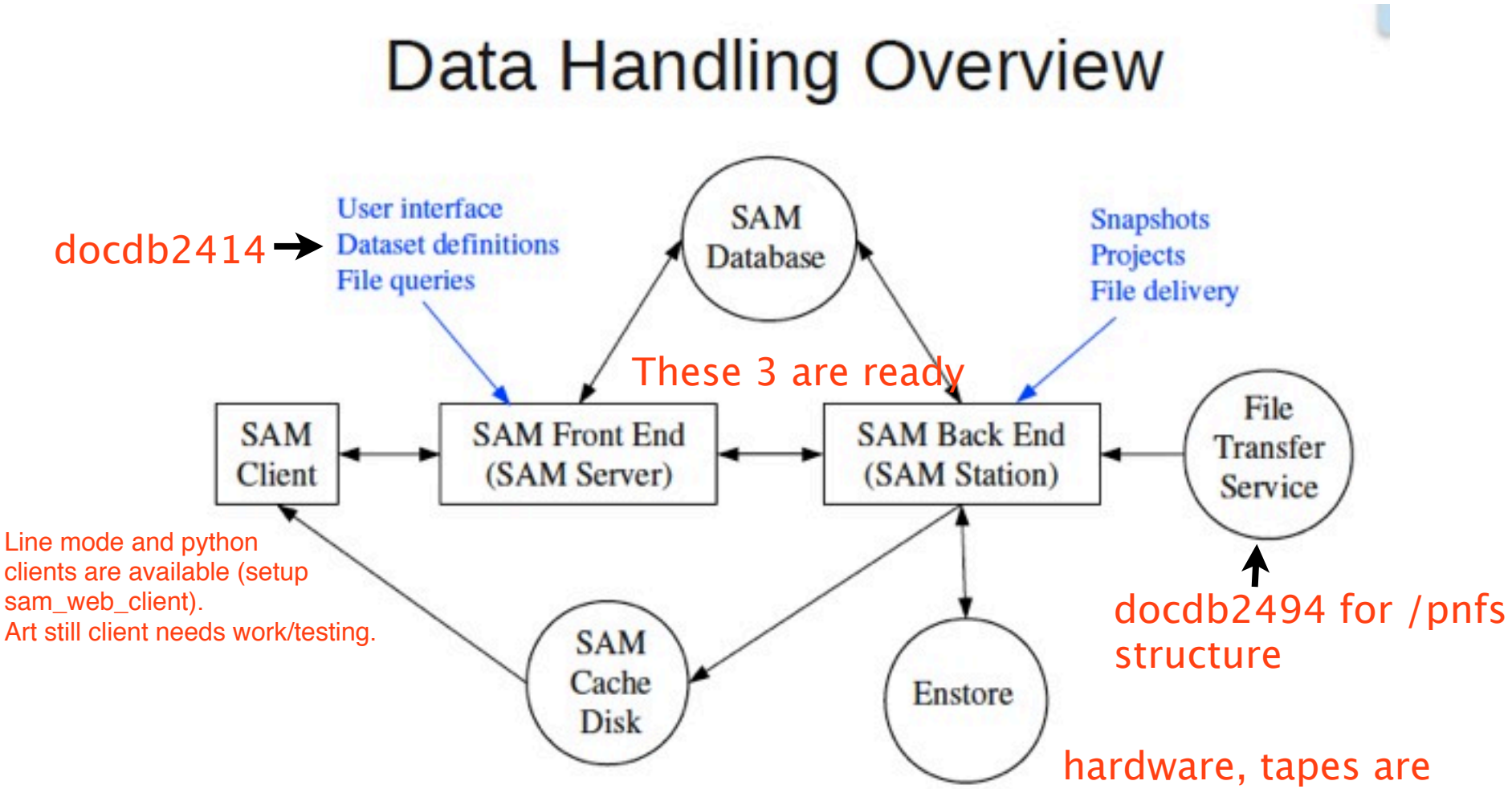
Outline

- LArSoft infrastructure
- Accounts, Grid, ...
  - next MC challenge
  - suggestions for user submission
  - Quotas
- Recon/Sim News
  - Recon work
  - Next MC Challenge
- uBooNE Offline
  - Reminder of our own repository
- Offline org Chart

LArSoft Project

- Reminder: SCD is now (er, almost) in charge
  - Rick Snider forming team,
  - For next 1-N (N~3?) months, Brian and I still call meetings, do releases, I think ...
  - The main thing MicroBooNE stands to benefit from is a stable release model, and perhaps even our own branch. Standby.
- SAM/FTS/Enstore/IFdh Progress
  - Herb is bringing the IF into the modern SAM world
  - Working with ART and REX to allow proper meta-data labeling of all our MC and Data files.
  - Working with ART to get LArSoft to allow proper Enstore storage of data to tape/disk and File Transfer Service of files back to our LArSoft jobs.
  - **Writing modules/fcl files to do all this.**
  - See https://cdcvs.fnal.gov/redmine/documents/625

SAM, etc


Data Handling Overview
docdb2414
User interface
Dataset definitions
File queries
SAM Database
Snapshots
Projects
File delivery
These 3 are ready
SAM Client
SAM Front End (SAM Server)
SAM Back End (SAM Station)
File Transfer Service
Line mode and python clients are available (setup sam_web_client).
Art still client needs work/testing.
SAM Cache Disk
Enstore
docdb2494 for /pnfs structure
Not config'd yet.
hardware, tapes are available. We'll test upload MC challenge files soon.

SAM, contd

- The point is to: add metadata to the files and properly name them and to enable seamless Enstore use of files.

- Insofar as we're concerned we eventually want the lay-user to just run with the appropriately modified fcl file to just do all this under the hood.
  - We (Herb! and ART) are not quite there yet.
  - Production jobs for MC Challenge will do this first, in stages. Metadata will be attached for next challenge.

LArSoft minutiae

- Frozen Release request
  - Are working on it now
  - Genie2.8

- Plane Label integers different -- only a problem if not using S2012.12.17 to read Challenge files
  - makes at least histograms wrong if the dev release is used

Accounts

- New system to setup up new users on **uboonegpvm0N** has been streamlined a bit
  - Better/Easier now, right? Right?: http://computing.fnal.gov/xms/Services/Getting_Started
  - The new welcome message will announce your membership to microboone@fnal.gov and redmine-larsoftsvn and list other listservs/projects to which you can sign up.
    - ubooneoffline, uboonedaq redmine projects
    - larsoft@fnal.gov, larsofttracking@fnal.gov, microboone_analysis@fnal.gov, uboonedaq@fnal.com
    - larforum.org/uboone
    - skype::LArSoft chat group
    - larsoftcommit@fnal.gov, uboonecommit@fnal.gov, uboonedaqcommit@fnal.gov
    - http://dbweb0.fnal.gov:8080/ECL/uboone/E/index
  - (separately: email me/Gennadiy for uboonedaq machine accts.)

Kerberos is your first step

- There can be some difficulties matching ssh_config protocols/flags
- kerberos connection details:
  - http://www.usqcd.org/fnal/troubleshooting.html
    - TMI?!
- Submit Service Tickets! at http://computing.fnal.gov
  - "Generally, I'm having a problem"->Accounts for authentication
- Have your new collaborator ask us (Herb, Eric, Mike K, ... Andrzej)!

uBooNE Tools: Recon updates

- New LArSoft Reconstruction work
  - New FuzzyClustering – Ben C
  - CornerFinding/EndPointFinding – Wes
  - Optical Sim and FlashFinder Updates – Ben J

Ben C's stuff

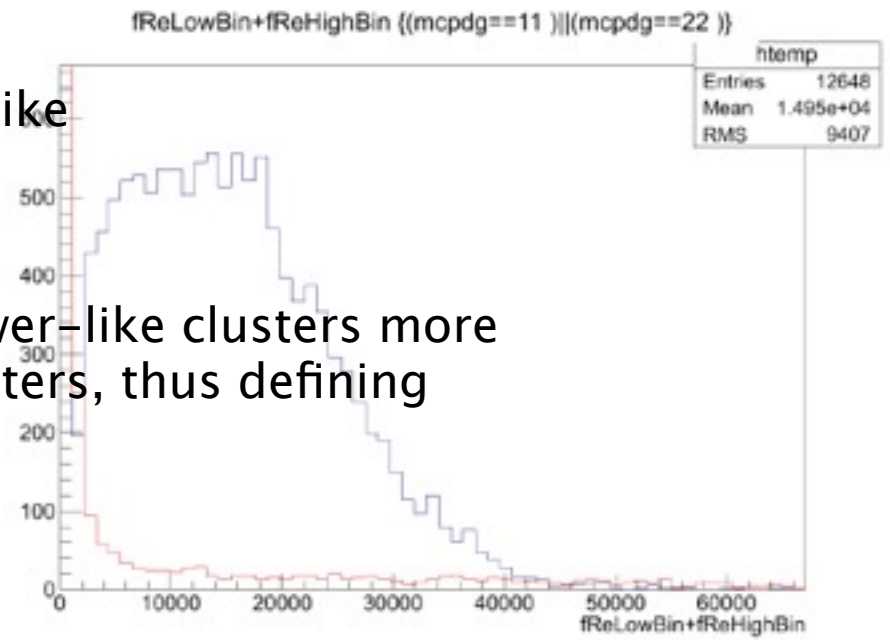
Shower likeness

- Around each line, sum up distance from a hit to the line, then divide by the charge of the hit, value will be large for showers

Shower likeness

Work with Andrzej
to separate track-like and shower-like clusters.

Will merge Hough lines within shower-like clusters more leniently than within track-like clusters, thus defining resulting cluster differently.



Purity and Efficiency

- Evaluated using the product of efficiency and purity found in the BackTracker

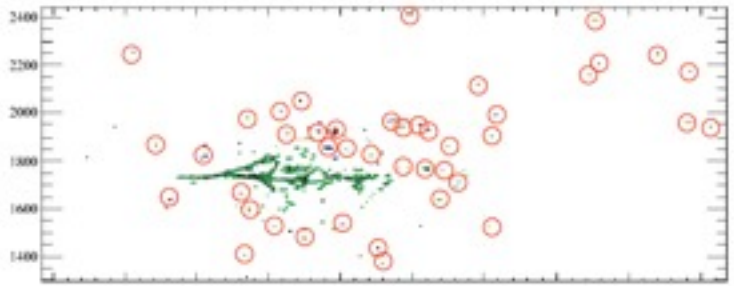
$$\text{Purity} = \frac{\#\text{hits from trackID in cluster}}{\text{total \# hits in cluster examined}}$$

$$\text{Efficiency} = \frac{\#\text{hits from trackID in cluster}}{\text{total \# hits for that trackID}}$$

Select values for trackIDs having the highest purity

- Comparison is tricky since Fuzzy Clustering and DBSCAN/Hough/Line Merger give different products
- Looked at 1,000 CCQE events

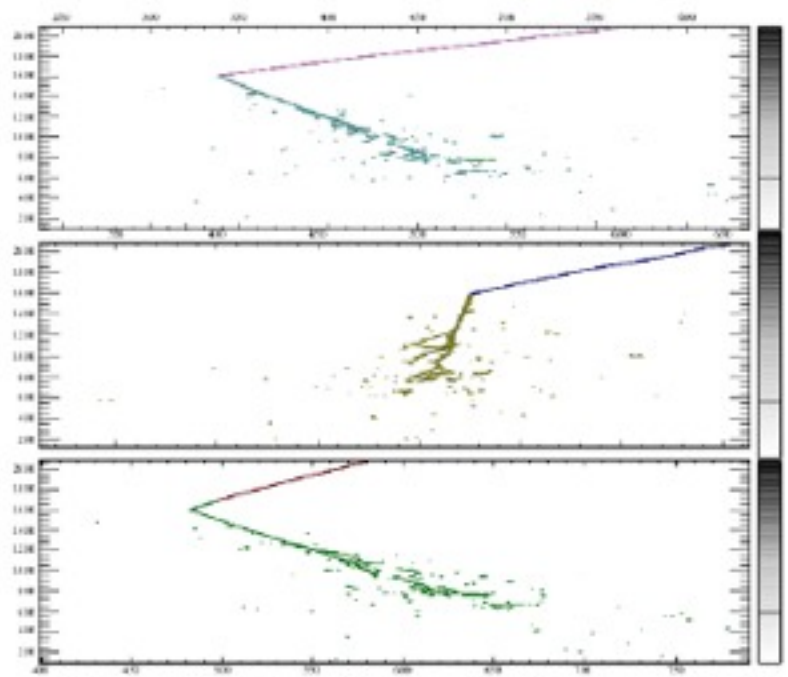
An example with DBSCAN

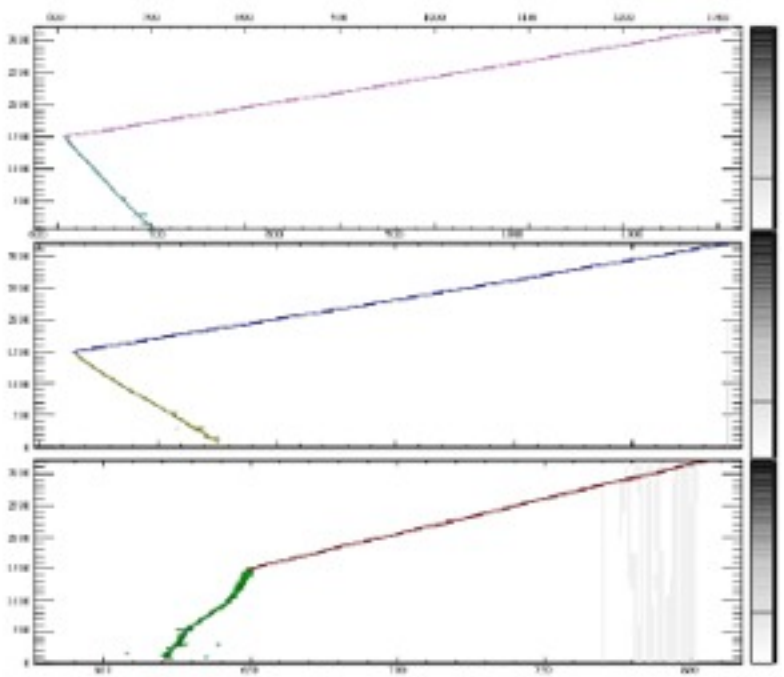


- We expect a lot of high purity clusters, lots of little clusters with only the

usters

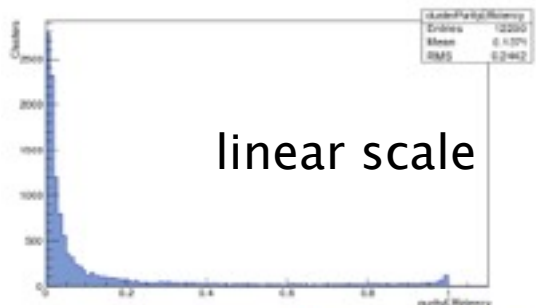
Fuzzy clustering

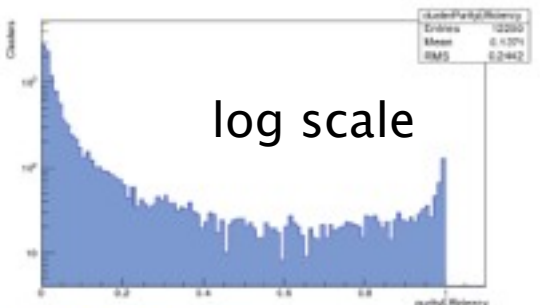




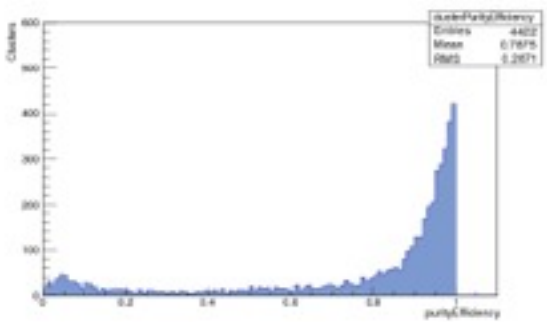
eff * purity

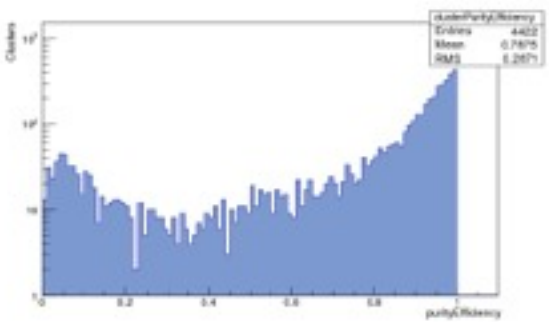
dBScan+Line Merger





Fuzzy clustering





Wes's stuff:

IMAGE PROCESSING: STRUCTURE TENSOR

Harris-Stephens is an image-processing technique based on locating large changes in "intensity"

- Take digitized 2D-image
- Create partial derivative images for both dimensions
  - I'll call the directions *x* and *y*
- Construct "structure tensor" over some local neighborhood of pixels

$$A = \sum_{u,v} \begin{bmatrix} \left(\frac{\partial I}{\partial x}\right)^2 & \left(\frac{\partial I}{\partial x}\right)\left(\frac{\partial I}{\partial y}\right) \\ \left(\frac{\partial I}{\partial x}\right)\left(\frac{\partial I}{\partial y}\right) & \left(\frac{\partial I}{\partial y}\right)^2 \end{bmatrix}$$

an Analog to old Harris vertex finder (which used Hits).

Analyze the eigenvalues of A

- The eigenvalues tell you how fast things are changing, and in how many dimensions

- $\lambda_1 \sim \lambda_2 \sim 0$
  - No big changes – things look similar in every direction
- $\lambda_1 >> \lambda_2 \sim 0$
  - See large directions along one direction (given by the eigenvector…), and not so much along the other
  - This would be an edge
- $\lambda_1 \sim \lambda_2 >> 0$
  - Big changes in intensity along all directions – a corner!
    - Or an endpoint, or a singularity

How to assign a Score to Each Pixel

Harris-Stephens

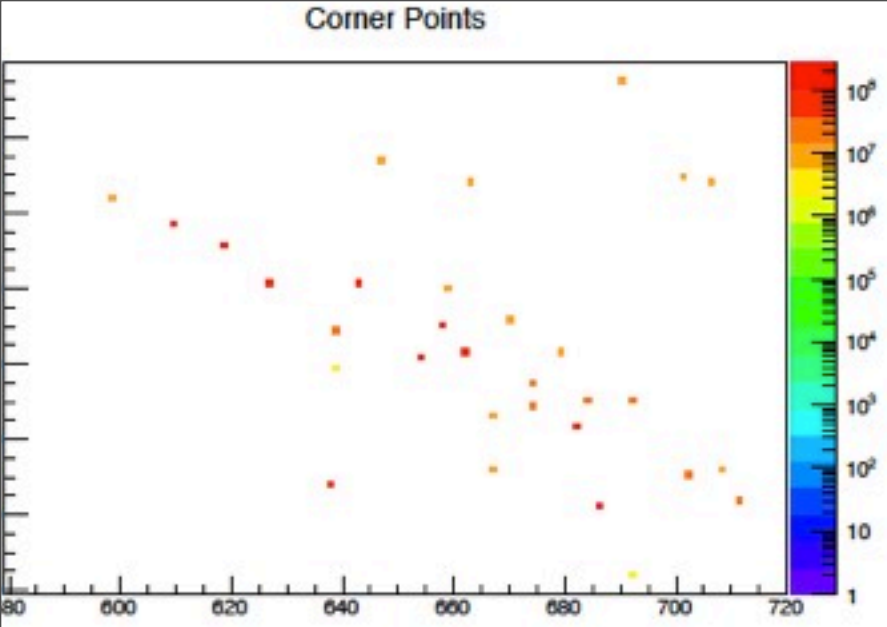
- $\det(\mathbf{A}) - \kappa \cdot \mathrm{tr}(\mathbf{A})^2 = (\lambda_1\lambda_2) - \kappa\,(\lambda_1+\lambda_2)^2$
  - κ is just empirically determined; typically ~ 0.05
  - For noise-space, score ~ 0; for edges, score < 0, and for corners, score > 0

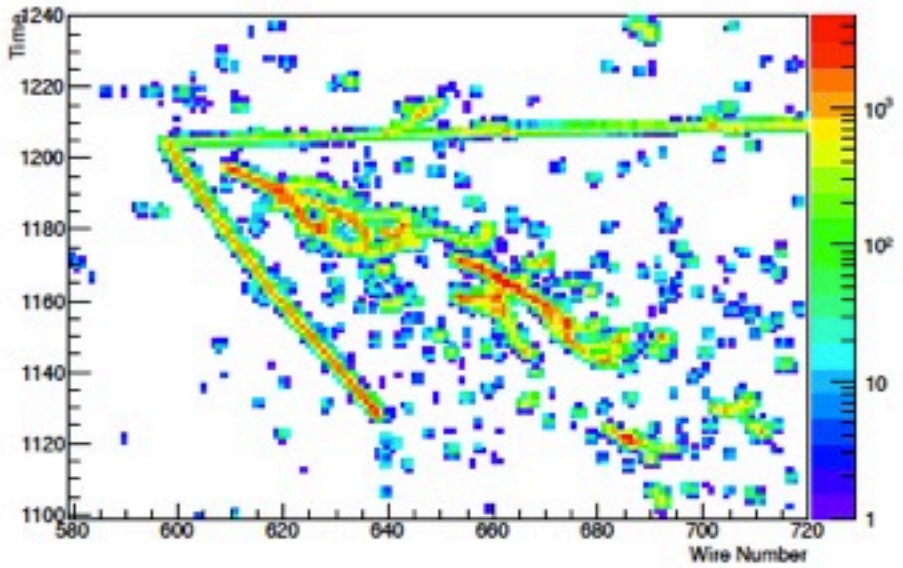
Noble (?)

- $\det(A) / \mathrm{tr}(\mathbf{A}) = (\lambda_1\lambda_2) / (\lambda_1+\lambda_2)$
  - For noise, score ~ 0; for edges, score is smallest eigenvalue; and, for corners, score ~ eigenvalue magnitude

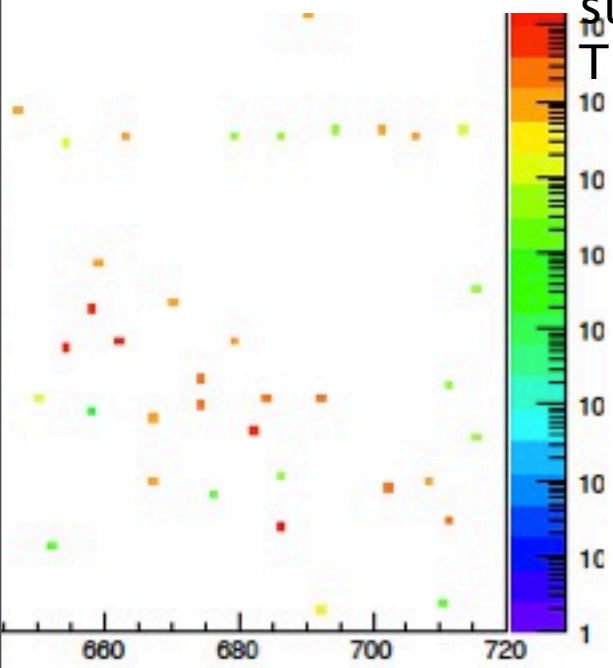
Shi-Tomasi

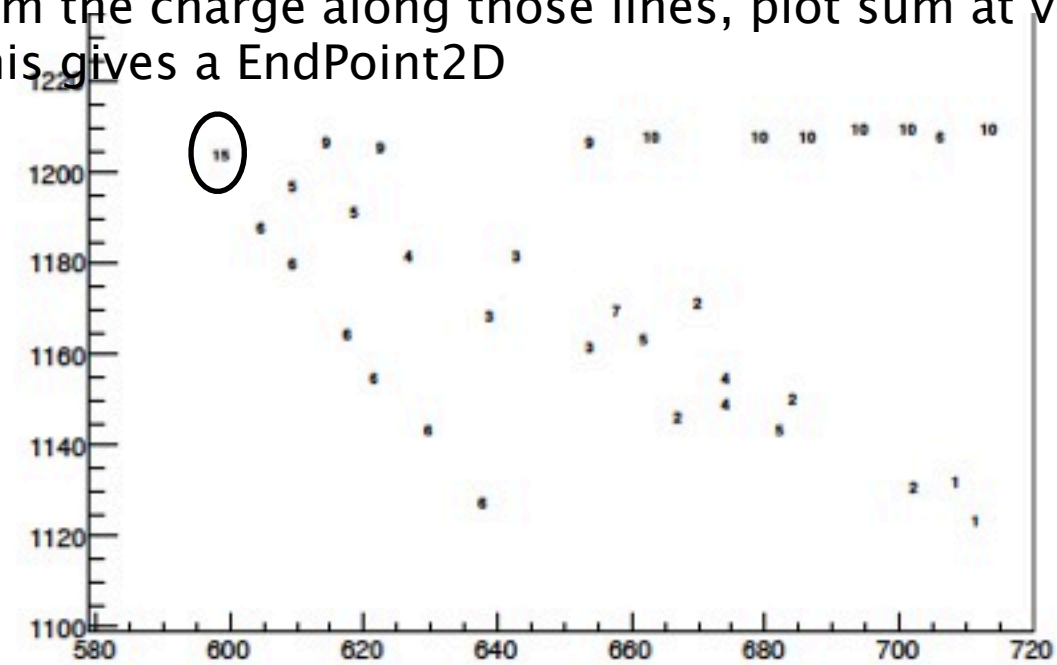
- $\min(\lambda_1, \lambda_2)$
  - More computationally intensive than the others…



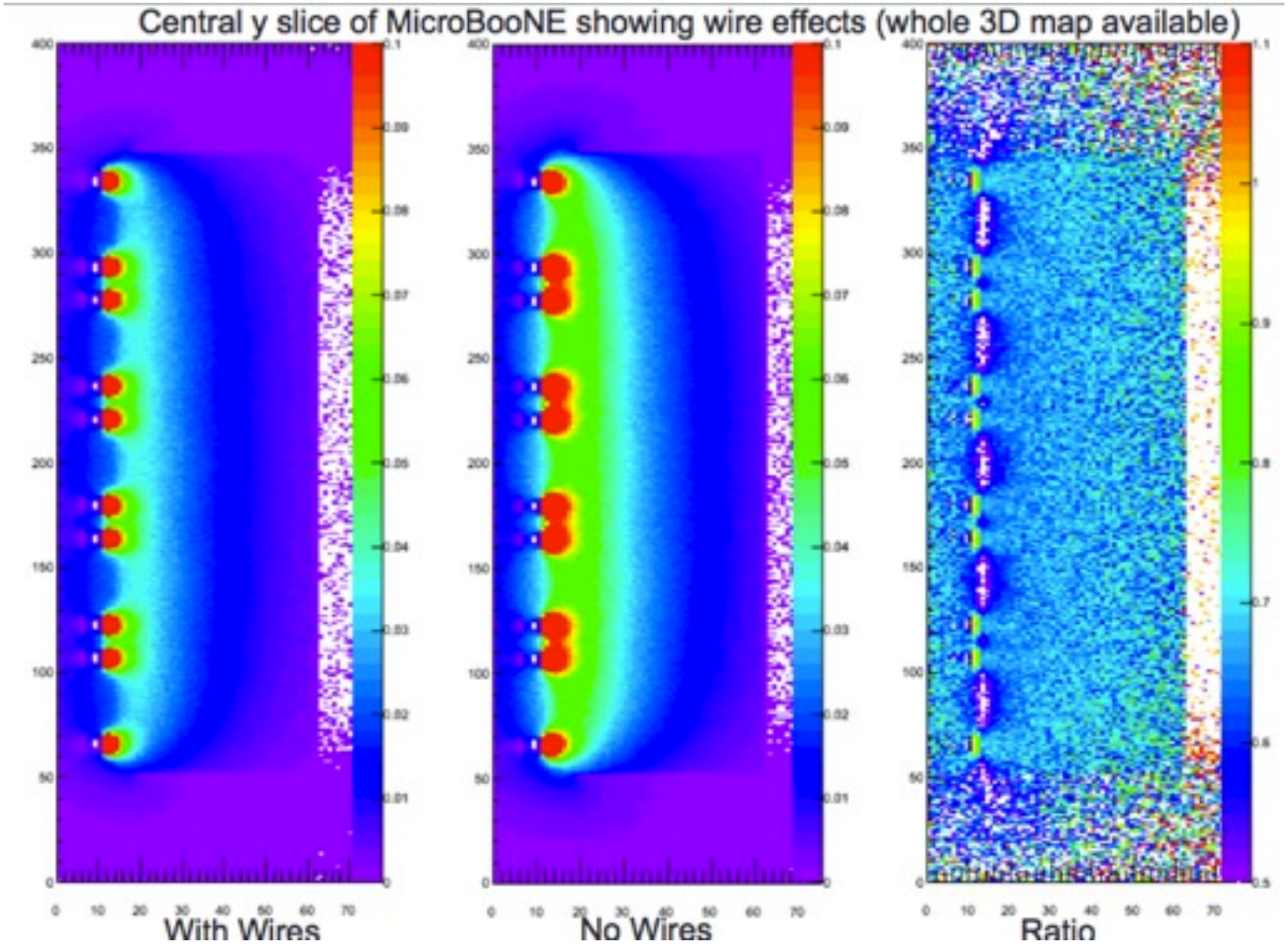


Draw Lines connecting locations of high Noble scores: sum the charge along those lines, plot sum at vertices. This gives a EndPoint2D

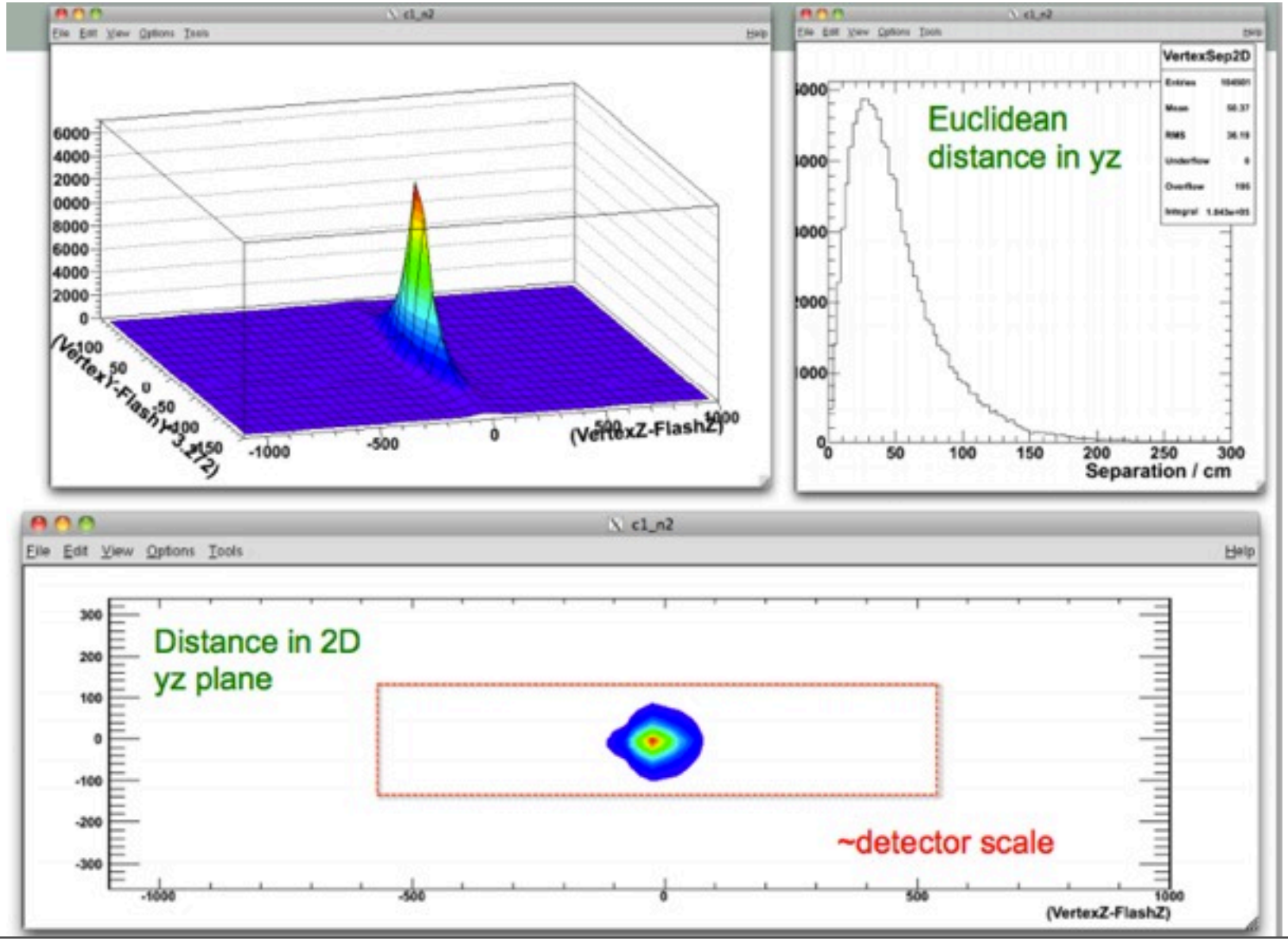




Building photon Look-up library with, w.o. wires



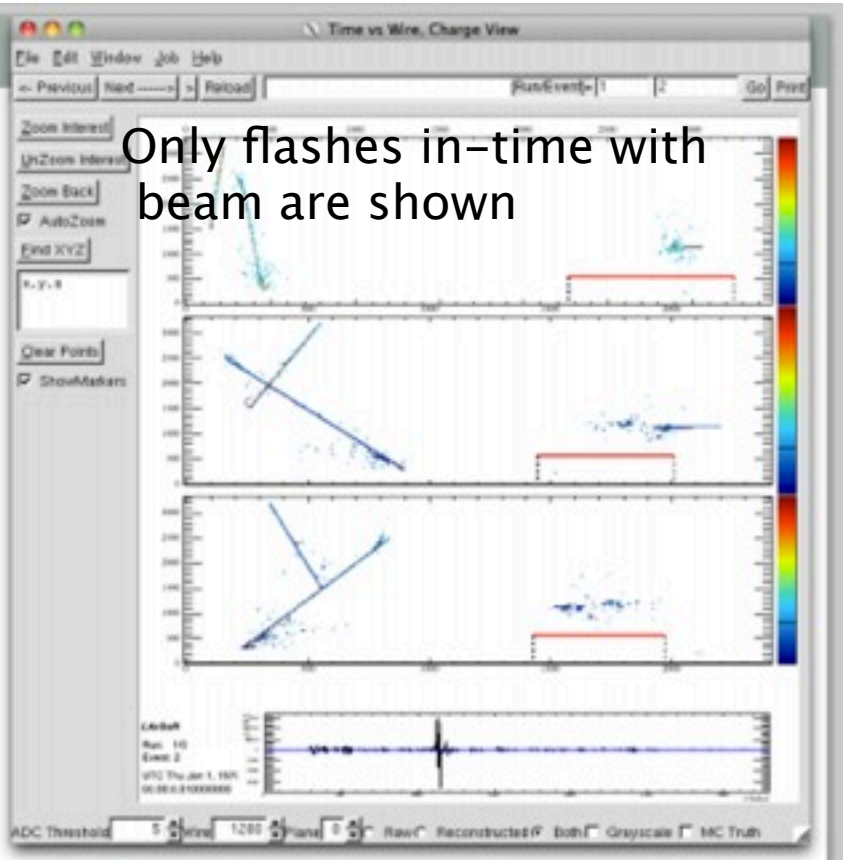
Flash finder resolution



with cosmic overlay

The FlashFinder:

- 1. Invert bipolar pulses to unipolar
- 2. Sum all unipolar pulses into a broad binned "superpulse"
- 3. Look for isolated spikes in this superpulse.
- 4. Each spike is a starting point for looking for a flash. Go back to each individual PMT waveform and look for the peak in this window.
- 5. For each channel, count the charge in some region around the peak, normd to 1PE
- 6. Combine all this information into a global object.



Flash properties:
Time,
TotalPE,
PE(pmt),
Center and width in yz,
Center and width in uv,

Next MC Challenge

- Did this once in Aug-2012.
- Want to do it again with ...
  - Optical Sim and FlashFinder Reco in
  - FuzzyClustering swap-in for dBCluster?
  - Cosmic timing in 3 windows - overlaid on some samples
    - Had x0.5 too few cosmics last time.
- should have disk space now
- **New MC Challenge end of May**
  - invitation for requests for samples
- Will attach metadata. Will not use SAM to store files away. ART-SAM interface not ready.

Our own repository

- ubooneoffline on redmine: ask Herb/me to be members
  - Analysis code
  - Bootlegged stuff
  - ups products: ubtools, ubfcl, ubxml, larsoft init
  - Auxiliary Geometries are candidates!
    - (Seligman's granite block at Nevis, but NOT Tia's NMSU veto system, it would seem. Really? What?)
  - **new** docs, users directories!
    - Please don't insert lotsa big images, etc
  - A community repository to complement a wide variety of work that's ongoing inside MicroBooNE
    - No code vetting
    - This area has no plans to conform to a bigger build system, unless people want that
- Note --- unrelated to this repository !!! -- 4 days bkup of /uboone/app/users/.snapshot. History to be extended, I think.

Canonical Interactive Setup

- **MicroBooNE offline code is largely a LArSoft enterprise.**

- You could put these four commands in an alias and hand-execute that alias each login. (There are .csh versions too.)

  - `<uboonegpvm02.fnal.gov> . /grid/fermiapp/products/uboone/etc/setups.sh`
  - `<uboonegpvm02.fnal.gov> . /grid/fermiapp/lbne/lar/code/larsoft/releases/development/setup/setup_larsoft_fnal.sh`
  - `<uboonegpvm02.fnal.gov> cd /uboone/app/users/XYZ/larsoft/testarea`
  - `<uboonegpvm02.fnal.gov> srt_setup -a`

Batch Submission, Quotas, etc

- Guidance
  - submit (jobsub)
  - condor_*.sh examples
    - This file has the lar -c yourscript.fcl line in it.
    - And the crucial ifdh cp
      - You **must** use ifdh cp to move files from your grid nodes back to /uboone/app,data/users !!!!!!
  - Using the latest versions, per https://cdcvs.fnal.gov/redmine/projects/ubooneoffline/wiki/Batch_Tools, at:
    - /grid/fermiapp/products/uboone/ubtools/devel/bin/submit_lar.sh
    - /grid/fermiapp/products/uboone/ubtools/devel/bin/condor_lar.sh
    - /grid/fermiapp/products/uboone/ubfcl/devel/sim,reco,evd/*.fcl
- Quotas coming soon
  - /uboone/data/users/XYZ 1.5 TB
  - /uboone/app/users/XYZ 0.3 TB
  - /uboone/data/randd,oscn,xsec/ 10 TB each

Org Chart

- HackDay (Wes's talk), Microboone_analysis@fnal.gov, fora, ...
  - Lots of new offline code enthusiasm!
- We thus have a new proposed Org Chart, with rather exhaustively detailed job descriptions.
  - Simulations
  - Reconstruction
  - MC Management
  - Data Management
  - dB Tools
  - Software Tools
  - EventDisplay
- Deliverables/milestones under construction for each category

Thoughts from the Naugehyde swivel chairs of the Tools Conveners:

Summer Student Work

- "Adopt a topology" is the current meme
- All studies will stem from just one file (of all the many files we generate) MC challenge, in this scheme
  - BNB flux + Genie evts
- One could imagine writing filters to pick out the desired topologies at the MCTruth level. A good LArSoft intro.
- We propose (if it's our place to do so) Flavio's studies as a potential summer exercise for small groups of students with a LArSoft "savvy" mentor

LArSoft Tutorials, etc

- We had one last month
  - How to run jobs, set up one's environment
  - very chatty, informal
  - no C++, so far.
  - I think it was useful
  - I heard a request for another, more in-depth session
  - Should we organize a few-session tutorial for students?
- We of course still have the "Walk-Through Exercises" on the LArSoft wiki
  - This is a nice first set of homework to start off a new MicroBooNE User

Bkup

Regarding the tasks and responsibilities listed below, we do not expect one person to take on all of the listed tasks him/herself (some tasks, yes). Rather, the leader's job is to monitor progress and coordiante. Also, we do not expect that candidate leaders will necessarily have all of the technical skills and knowledge to do the listed tasks at the outset (but we expect that people are able to learn).

I. Simulation leader.

Reponsible for standard simulation algorithms and job files. Algorithms generally fall under the following three categories.

a) Generators.
b) Geant simulation.
c) Detector simulation.

Reponsibilities:

a) Become familiar with all generators in use or being considered or developed. Integrate new generators into larsoft as needed.
b) Become familiar (if not expert) with how the geant simulation is organized and configured (e.g. physics lists).
c) Maintain simulation job files.
d) Maintain microboone geometries (add new elements as needed).
e) Develop simulation verification/test suite.

Some specific high priority tasks:

a) Improved noise simulation.
b) Improved electronics response.
c) Improved (recalculated) field reaponse, including different field response for first two induction planes.
d) PMT readout and trigger simulation.

II. Reconstruction leader.

Responsible for reconstruction algorithms and job files. Algorithm categories are not as well-defined as simulation, but include the following:

a) Signal shaping (deconvolution).
b) Hit finding.
c) 2D Cluster finding.
d) 2D and 3D vertex finding.
e) 2D and 3D track reconstruction.
f) 2D and 3D shower reconstruction.
g) Optical reconstruction.

Responsibilities:

a) Become familiar with all of the standard reconstruction algorithms.
b) Monitor progress in each of the above categories, with a view to determining which algorithm(s) may be falling behind, or which need more effort.
c) Integrate all of the algorithms that make up "standard reco." Maintain standard reco job files.
d) Develop a verification/test suite for the reconstruction program.

III. MC Production Leader

Responsibilities:

a) Develop MC production and bookkeeping systems.
b) Document all standard MC samples (e.g. on a web page).
c) Generate MC.

IV. Data Production Leader

Reponsibilities:

a) Devloping data production and bookkeeping systems for fast
turnaround data and physics analysis data (if different).
b) Document all standard raw and reconstructed data samples (e.g.
on a web page).
c) Develop and execute procedure for converting binary raw data to
art format.
d) Reconstruct data.

V. Database Leader.

Responsibilities:

a) Identify which databases need to be developed.
b) Design database schemas (database design).
c) Develop methods for populating and extracting data from databases (database applications).
d) Develop a method for serving database data to farms and remote sites (database servers).

In many cases, technical details of the above tasks may have to be handled by experts in SCD and CCD. The database leader should take on the responsibility and decision-making authority to make the above tasks happen.

VI. Software Tools Leader.

Tools categories:

a) Grid job submission (jobsub_tools).
b) Data handling (art and ifdh).
c) Other

Responsibilities:

a) Become familiar with external tools in use and being developed.
b) Developing microboone-specific layered tools and scripts.
c) Develooping and maintaining the "microboone computing environment," at Fermilab and at remote sites (including maintaining microboone ups products area).
d) Manage releationship with larsoft (e.g. make sure microboone environment stays compatible with larsoft environment).

Some high priority tasks:

a) Design sam metadata.
b) Configure enstore and file transfer service.
c) Integrate sam into larsoft and job submission tools.

VII. Event Display Leader.

Tools categories:

a) LArSoft native EVD
b) Argo EVD
c) Other

Responsibilities:

a) Become familiar with the two EVDs
b) Develop microboone-specific configurations for the LArSoft EVD.
c) Develop and Maintain LArSoft's EventDisplay code and be conversant
in Nusoft's EventDisplayBase's code
d) Bring Argo's codebase to some accessible
location, as needed. Learn about and manage its build and distribution system.
Make contributions to its code.

Some high priority tasks:

a) Run Argo, Run EventDisplay, learn when one is appropriate to which MicroBooNE needs.
b) Advocate with other experts for development of each, as needed
c) Educate collaboration on use of each event display tool.
d) Understand Online/Nearline's use of each.
e) Understand Shifter's use of each